

# How to implement and verify FAIR principles ?

Alban Gaignard

CNRS, institut du thorax, Nantes, France

November 24, 2024

ANF "Workflow et reproductibilité" 2025, Paris

# A practical overview on FAIR assessment, in 2 hours ...

## ▶ Quick intro on FAIR principles

## ▶ Knowledge graphs and semantic metadata

-  What are Knowledge Graphs, Computational Ontologies ?
-  How to create Knowledge Graphs ? (team work, "pen & paper")
-  How to write machine-readable semantic metadata ? (demo, python code)
-  How to query knowledge graphs with SPARQL (demo, python code)

## ▶ Tools and resources for increased FAIRness

-  FAIR-Checker
-  Tools and workflow registries
-  Using FAIR-Checker and interpreting the FAIR assessment results
-  Using the FAIR-Checker API and processing multiple resources



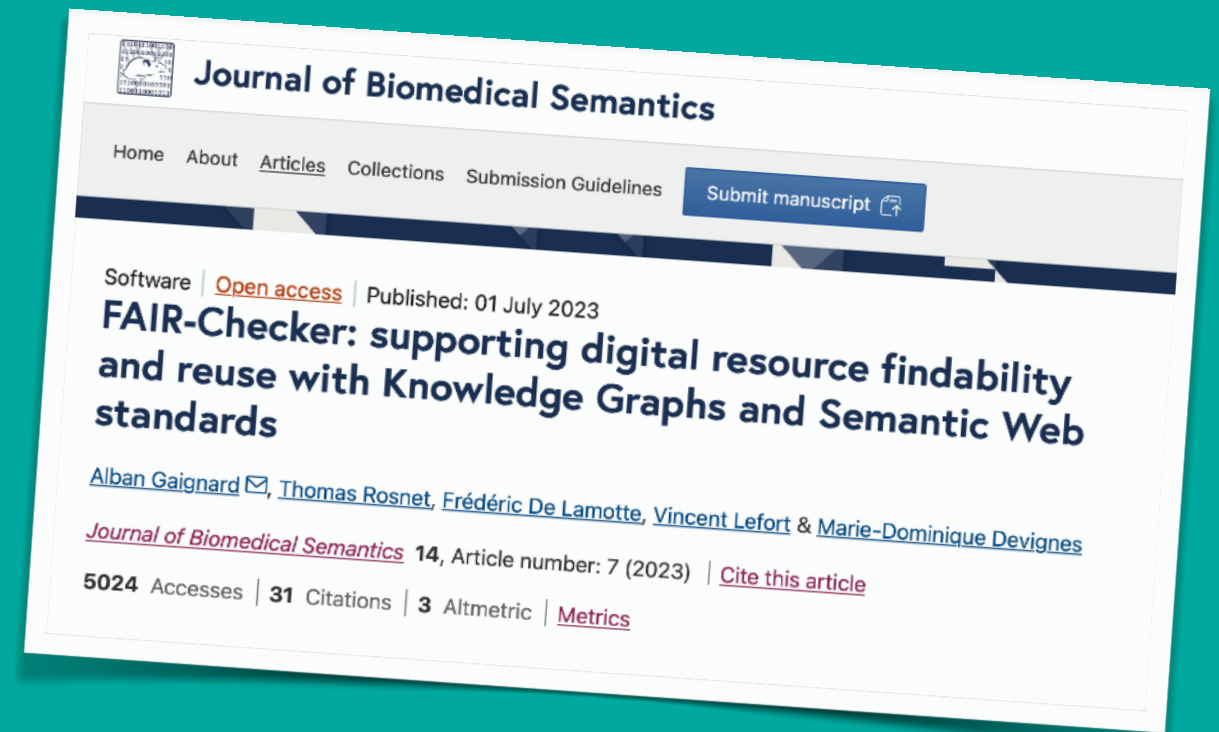
Marie-Dominique Devignes



Thomas Rosnet



Frédéric de Lamotte



# How can we evaluate that research data are FAIR enough ?

# FAIR principles require tooling



Australian Research Data Commons

*Web URIs*

*HTTP, RDF/SPARQL*

*Domain ontologies*

## FAIR principles

- ▶ critical for open & reproducible sciences
- ▶ result in many guidelines
- ▶ technology agnostic guidelines

How to implement the principles ...  
... and go beyond checklists ?

**Resource provider  
and developers  
need **help** and **tooling**.**

<https://www.go-fair.org/fair-principles>

<https://www.nature.com/articles/sdata201618>

# Usage scenarios



**Data producers**

**Make**



- Dataset
- Training
- Software tools
- etc ...

**discoverable !**

**Where to publish ?**

**Which registry ?**

**Does it provide metadata ?**

**Is it enough to be FAIR ?**



**Registry developers**

**Make**



- Dataverse
- Bio.tools
- Zenodo,  
Pubmed ...

**more FAIR !**

**Improve metadata quality ?**

**Community specific  
standards ?**

**Which technology ?**

# Why a (nother) tool ?

## Assumptions

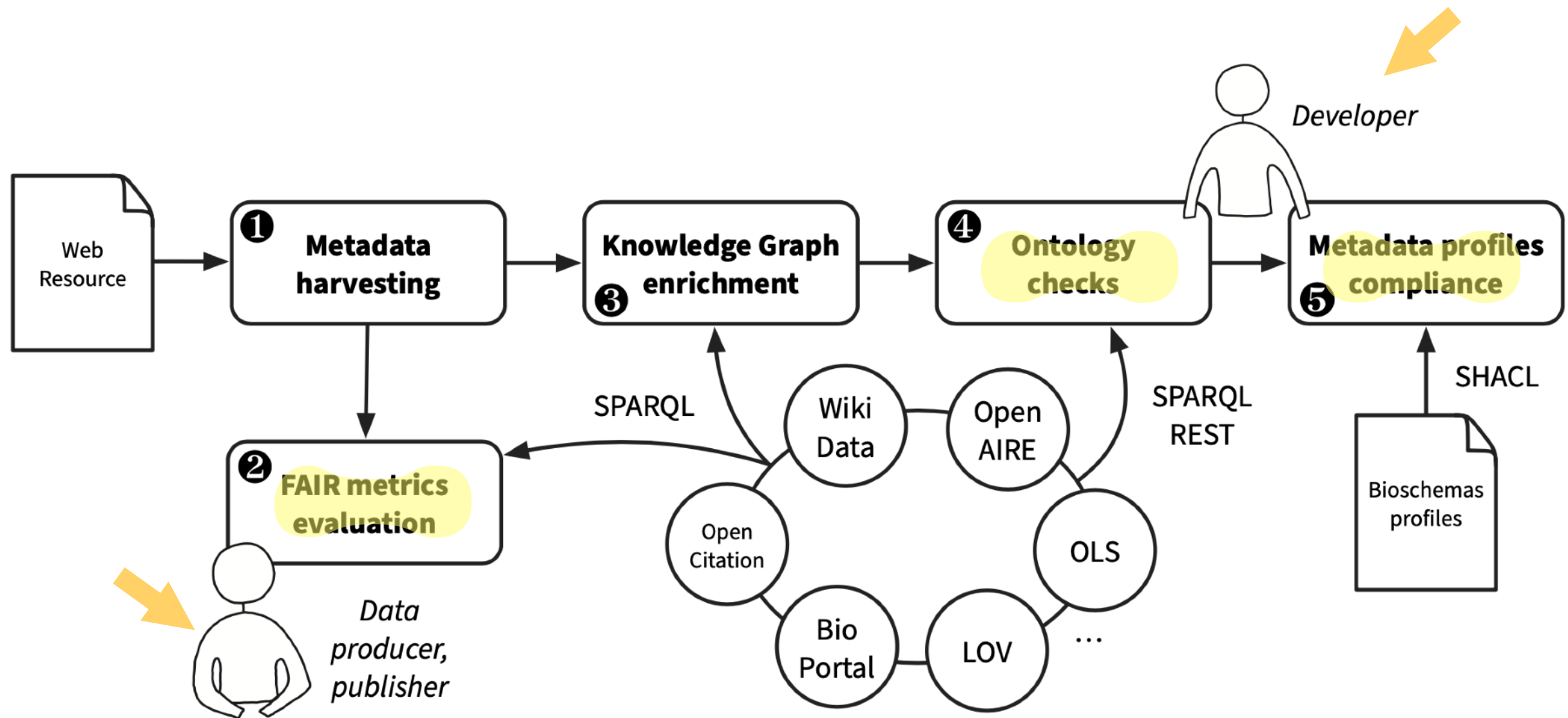
- ▶ "Linked Data" and Semantic Web technologies are key in most of the FAIR principles (especially F, I, and R)  
... but technical skills are needed.

## Objectives

- ▶ Provide a web interface for resource providers to **evaluate FAIR metrics** and **make progress on FAIRification** (iterative testing)
- ▶ Provide additional tools for developers Leverage semantic web technologies (RDF, SPARQL, SHACL) to **enhance the quality of metadata**
- ▶ Share implementations, and contribute to larger FAIR evaluation initiatives



# Approach



**Figure 1** Gathering, enriching and analyzing semantic web annotations in line with FAIR principles.

# ① Checking metadata

# Checking metadata



Semantic web technologies

- RDF
- SPARQL

<https://fairplus.github.io/the-fair-cookbook/content/home.html>

<https://rdmkit.elixir-europe.org>

# What is evaluated ?

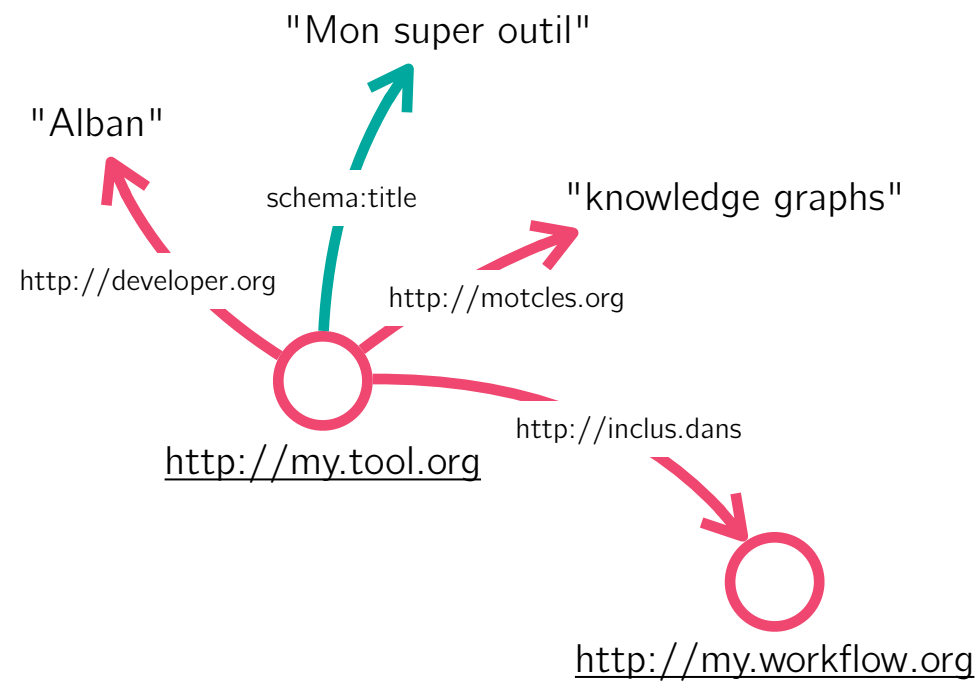
| Findability<br>F1B, F2                                                                                                                                    | Accessibility<br>A1.2                            | Reuse (licenses)<br>R1.1                                                                                                | Reuse (provenance)<br>R1.2                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| dct:identifier<br>schema:identifier<br>dct:title<br>dct:description<br>dcat:accessURL<br>dcat:downloadURL<br>dcat:endpointDescription<br>dcat:endpointURL | odrl:hasPolicy<br>dct:rights<br>dct:accessRights | schema:license<br>dct:license<br>doap:license<br>dbo:license<br>cc:license<br>xhv:license<br>sto:license<br>nie:license | prov:wasGeneratedBy<br>prov:wasDerivedFrom<br>prov:wasAttributedTo<br>prov:used<br>prov:wasInformedBy<br>prov:wasAssociatedWith<br>prov:startedAtTime<br>prov:endedAtTime<br>dct:hasVersion<br>dct:isVersionOf<br>dct:creator<br>dct:contributor<br>dct:publisher<br>pav:hasVersion<br>pav:version<br>pav:hasCurrentVersion<br>pav:createdBy<br>pav:authoredBy<br>pav:retrievedFrom<br>pav:importedFrom<br>pav:createdWith<br>pav:retrievedBy<br>pav:importedBy<br>pav:curatedBy<br>pav:createdAt<br>pav:previousVersion<br>schema:creator<br>schema:author<br>schema:publisher<br>schema:provider<br>schema:funder |

DC-Terms  
DCAT  
Schema.org

+ ODRL + DOAP,  
DBO, CC ...  
  
+ PROV-O, PAV

**Table 2** Summary of the selected ontology properties relevant to assess three specific FAIR principles in *FAIR-Checker*

# Example "http://my.tool.org"

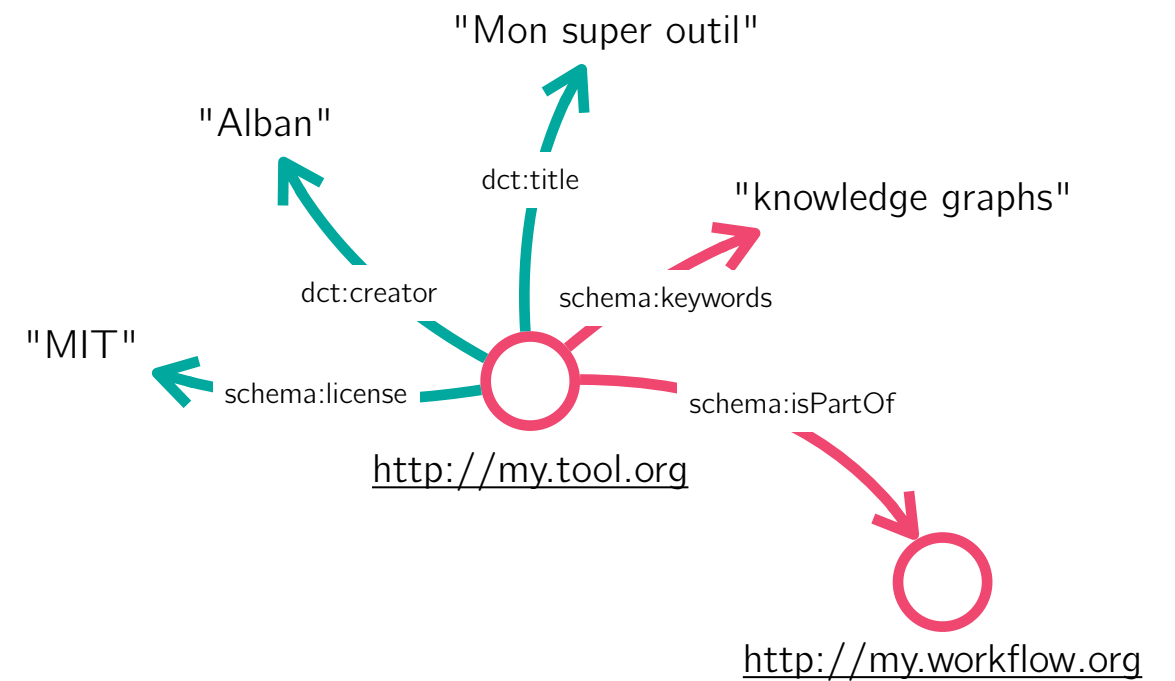


F1B, F2 → ✓

A1.2 → 🚫

R1.1 → 🚫

R1.2 → 🚫



F1B, F2 → ✓

A1.2 → 🚫

R1.1 → ✓

R1.2 → ✓

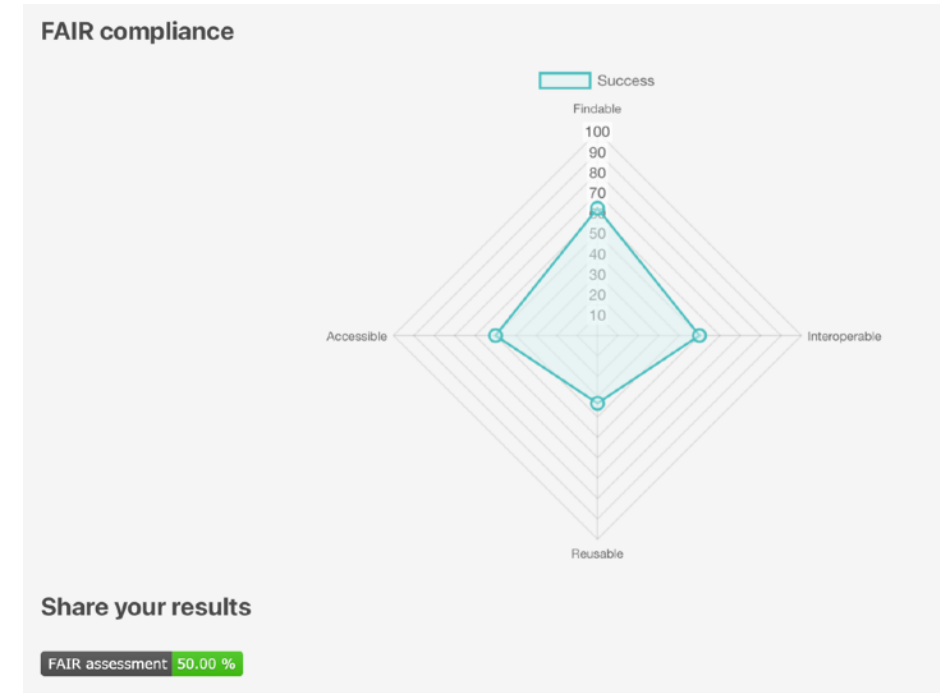
**Resource identifier (URL/DOI)**

   [All metrics](#)

Valid URL/DOI - The input contains the following DOIs that you can also test:  10.3205/09dgnc137

 [Clean results](#)

[Dataset Dataverse](#) [Workflow](#) [Publication Datacite](#) [Dataset](#) [Tool](#)



|                                                |                       |                                |                                                                                                                                                                                                |                                                                                       |
|------------------------------------------------|-----------------------|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| F2B: Shared vocabularies for metadata          | <a href="#">Check</a> | FAIR principle F2B <b>2/2</b>  |                                                                                                                                                                                                |    |
| A1.1: Open resolution protocol                 | <a href="#">Check</a> | FAIR principle A1.1 <b>2/2</b> |                                                                                                                                                                                                |   |
| A1.2: Authorisation procedure or access rights | <a href="#">Check</a> | FAIR principle A1.2 <b>0/2</b> | You should describe the access policy in metadata by using at least one of the                                                                                                                 |  |
|                                                |                       |                                | <a href="#">Read more ▾</a>                                                                                                                                                                    |                                                                                       |
| I1: Machine readable format                    | <a href="#">Check</a> | FAIR principle I1 <b>1/2</b>   | You should provide discoverability oriented metadata with one of the following properties: dct:title dct:description dcat:accessURL dcat:downloadURL dcat:endpointDescription dcat:endpointURL |  |
|                                                |                       |                                | <a href="#">Read less ▲</a>                                                                                                                                                                    |                                                                                       |
| I2: Use shared ontologies                      | <a href="#">Check</a> | FAIR principle I2 <b>2/2</b>   |                                                                                                                                                                                                |  |
| I3: External links                             | <a href="#">Check</a> | FAIR principle I3 <b>0/2</b>   | You should enrich your metadata with more diversified external links. Here we did not                                                                                                          |  |
|                                                |                       |                                | <a href="#">Read more ▾</a>                                                                                                                                                                    |                                                                                       |
| R1.1: Metadata includes license                | <a href="#">Check</a> | FAIR principle R1.1 <b>0/2</b> | You should include information about licence in your metadata using one of the                                                                                                                 |  |
|                                                |                       |                                | <a href="#">Read more ▾</a>                                                                                                                                                                    |                                                                                       |

1. Provide resource URL
2. Retrieve metadata
3. Evaluate relevant metadata fields (ontology classes and properties)
4. Propose **relevant recommendations** from FAIR Cookbook and RDMkit
5. Improve metadata and **try** again ...

# Hands-on session: playing with FAIR-Checker

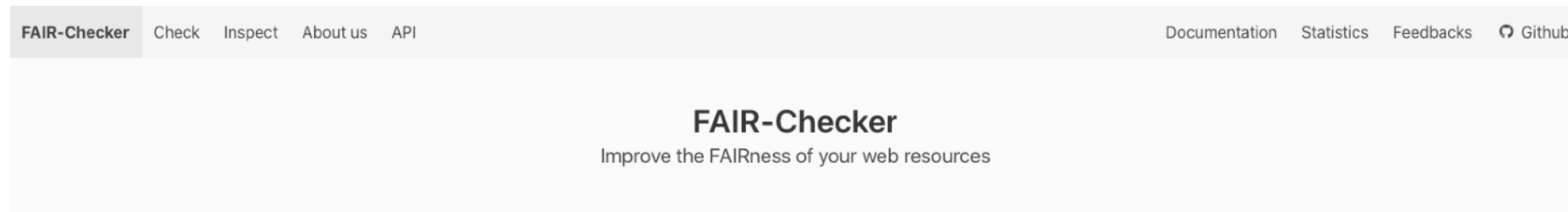
## Question #5

1. Evaluate the FAIRness of
  - a github repository:  
<https://github.com/IFB-ElixirFr/fair-checker>
  - a zenodo dataset:  
<https://doi.org/10.5281/zenodo.17681549>
  - a bioinformatics tool (Bio.Tools):  
<https://bio.tools/GrapeMine>
  - a bioinformatics workflow (NF-Core):  
<https://nf-co.re/scrnaseq/4.1.0/>
  - a bioinformatics workflow (WorkflowHub):  
<https://workflowhub.eu/workflows/1021>
2. Interpret the results

# Hands-on session: playing with FAIR<sub>Checker</sub>



<https://fair-checker.france-bioinformatique.fr/>



## Welcome

FAIR-Checker is a tool aimed at assessing FAIR principles and empowering data provider to enhance the quality of their digital resources.

Data providers and consumers can **check** how FAIR are web resources. Developers can explore and **inspect** metadata exposed in web resources.

Check ✓

Inspect 🔍



## Inspect: how it works ?

FAIR-Checker leverages semantic web technologies to check that metadata use standards and recognized ontologies or controlled vocabularies.

First, embedded semantic annotations are extracted from web pages, forming a minimal knowledge graph (Step 1). Then, this minimal knowledge graph is completed based on already deployed knowledge graphs: [Datacite](#), [OpenAire](#), [WikiData](#) (Step 2). Finally (Step 3), the resulting knowledge graph is tested to check that classes and properties are recognized through [Linked Open Vocabularies \(LOV\)](#), [Ontology Lookup Service \(OLS\)](#), or [Bioportal](#). The last verification (Step 4) consists in validating the resource metadata against [Bioschemas community profiles](#).

**Registries are instrumental** to  
improve FAIRness of digital artifacts ...

# Bio.tools registry



30617 tools

Explore ▾

Login

Sign-up

## GrapeMine (biotools:GrapeMine) 👍 ID Verified

<http://urgi.versailles.inra.fr/GrapeMine>

Available versions

1.0

Data management >

Genomics >

Proteomics >

✓ EDAM Topics

Legacy

LGPL-2.1

Free of charge

Database portal

Web service

Web application

Java

R

Python

JavaScript

Ruby

Perl



An integrated database for grapevine data

Sequence features >

Ontology data >



✓ EDAM Operations

Data retrieval >

Visualisation >

Query and retrieval >

Gene-set enrichment analysis >



✓ EDAM Data

Over-representation data >

Query script >

Sequence features >

✓ EDAM Data

### Documentation

<http://intermine.readthedocs.org/en/latest/web-services/>

API documentation

<https://flymine.readthedocs.io/en/latest/>

User manual

### Links

<https://registry.intermine.org>

Software catalogue

<http://intermine.org>

Other



# Bio.tools registry

```
<https://bio.tools/grapemine> a sc:SoftwareApplication ;
  ns2:input <https://bio.tools/grapemine/op_1/in_1>,
    <https://bio.tools/grapemine/op_1/in_2> ;
  ns2:output <https://bio.tools/grapemine/op_1/in_1>,
    <https://bio.tools/grapemine/op_1/in_2>,
    <https://bio.tools/grapemine/op_1/in_3> ;
  dct:conformsTo "https://bioschemas.org/profiles/ComputationalTool/1.0-RELEASE" ;
  sc:applicationSubCategory <http://edamontology.org/topic_0121>,
    <http://edamontology.org/topic_0622>,
    <http://edamontology.org/topic_3071> ;
  sc:citation <https://doi.org/10.1093/bioinformatics/bts577>,
    "pmcid:PMC3516146",
    "pubmed:23023984" ;
  sc:featureList <http://edamontology.org/operation_0224>,
    <http://edamontology.org/operation_0337>,
    <http://edamontology.org/operation_2422> ;
```

<https://bio.tools/GrapeMine#html>: 43 Triples

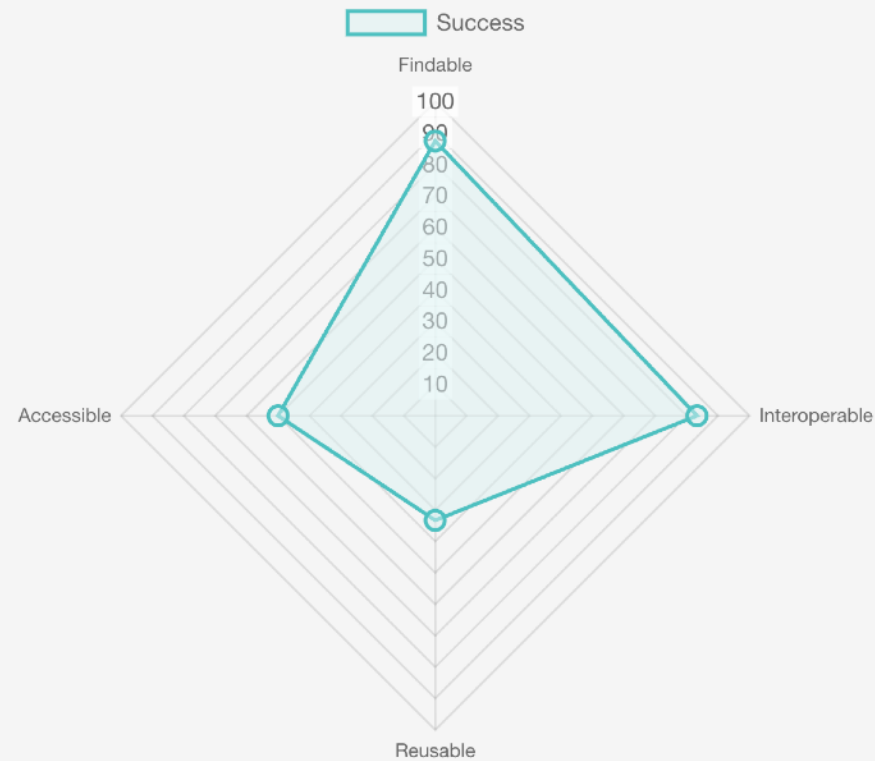
JSON-LD

Trig

→ 43 RDF triples retrieved by FAIR-Checker  
and used for FAIR assessment

# Bio.tools registry

## FAIR compliance



## Share your results

FAIR assessment 66.67 %

|                                    |       |                         |                                                                            |                             |  |
|------------------------------------|-------|-------------------------|----------------------------------------------------------------------------|-----------------------------|--|
| R1.1: Metadata includes license    | Check | FAIR principle R1.1 0/2 | You should include information about license in your metadata using one of | <a href="#">Read more ▾</a> |  |
| R1.2: Metadata includes provenance | Check | FAIR principle R1.2 0/2 | You should include information about provenance in your metadata using     | <a href="#">Read more ▾</a> |  |
| R1.3: Community standards          | Check | FAIR principle R1.3 2/2 |                                                                            |                             |  |

# WorkflowHub registry

 **nf-core/scrnaseq** 4.0.0 (latest) ▾

Overview

Files

Related items

Workflow Type: Nextflow



 GitHub Actions CI Status

nf-core linting

passing

CI tests

full size

DOI 10.5281/zenodo.3568187

unit tests

nf-test

nextflow DSL2 ≥24.04.2

run with conda

run with docker

run with singularity

Launch 🚀

Sequera Platform

slack

nf-core #scrnaseq

twitter @nf\_core

mastodon nf\_core

youtube nf-core

## Introduction

nf-core/scrnaseq is a bioinformatics best-practice analysis pipeline for processing 10x Genomics single-cell RNA-seq data.

This is a community effort in building a pipeline capable to support:

- SimpleAF(Alevin-Fry) + AlevinQC
- STARSolo
- Kallisto + BUStools
- Cellranger

[!IMPORTANT] Cellranger is a commercial tool from 10X Genomics Inc. and falls under the EULA from 10X Genomics Inc. The container provided for the CellRanger functionality in this pipeline has been built by the nf-core community and is therefore *not supported by 10X genomics* directly. We are in discussions with 10X on how to improve the user experience and licence situation for both us as a community

SEEK ID: <https://workflowhub.eu/workflows/1021?version=17>

## Version History

4.0.0 (latest)

Created 26th Mar 2025 at 09:57 by WorkflowHub Bot

 View on GitHub

 Download RO-Crate

📘 Creators and Submitter

Creators

Peter J Bailey, Bailey PJ, Alexander Peltzer, Botvinnik O, Olga Botvinnik, Marques de Almeida F, Peltzer A, Sturm G

Submitter

 WorkflowHub Bot

License

MIT License (MIT)

Activity

Views: 13191 Downloads: 7153

Created: 4th Jun 2024 at 11:33

Last updated: 26th Mar 2025 at 09:57

📘 Tags

10x-genomics 10xgenomics alevin bustools Cellranger kallisto

rna-seq single-cell star-solo

📘 Attributions

None

# WorkflowHub registry

Schema.org Documentation Schemas Validate About

<https://workflowhub.eu/workflows/1021>

Exécuter un nouveau test

あ

i

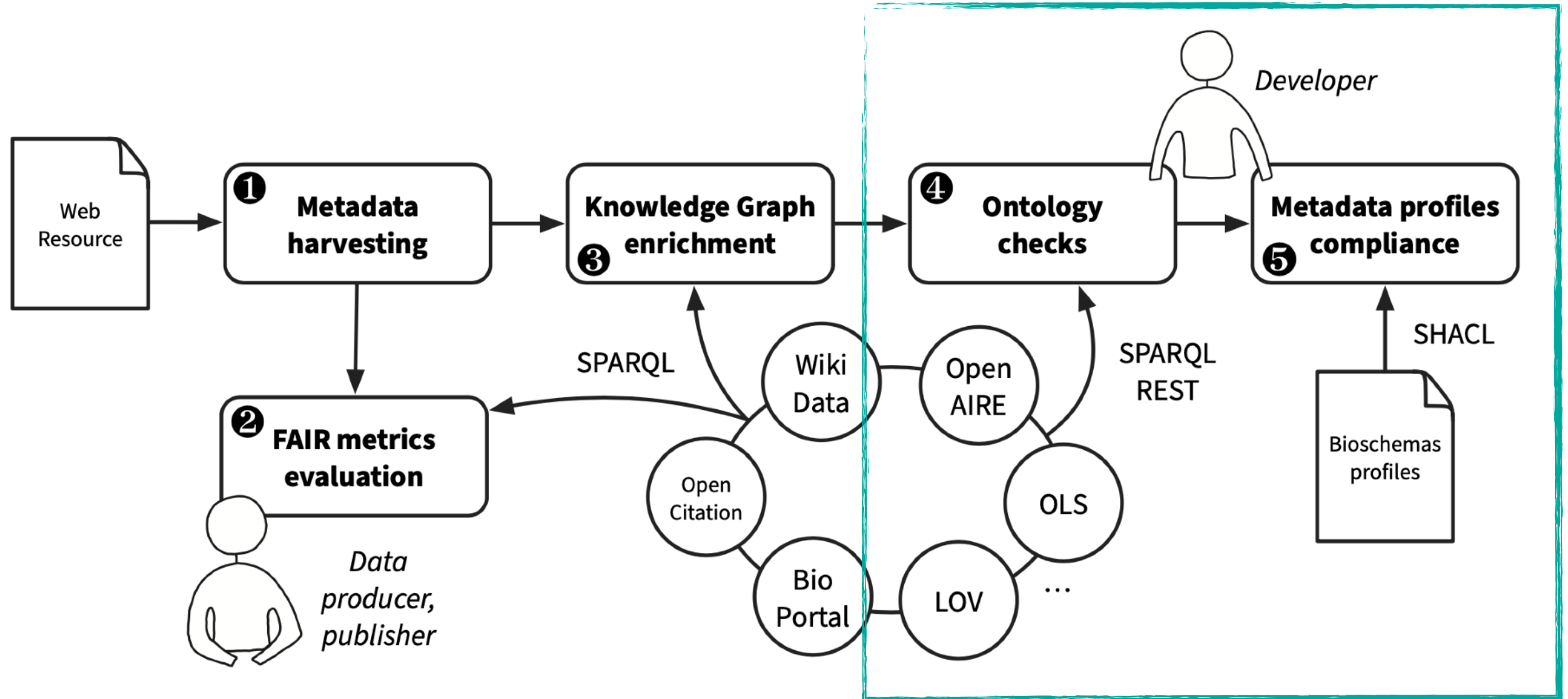
```
8 <meta name="csrf-param" content="authenticity_token" />
9 <meta name="csrf-token" content="3iMT2PTi5mzfM2t95rUKIz8Y2C8kSqwzkzc3R4sT0LCV-yIpobycXc5o6dkHjxagPyc
10 <script type="application/ld+json">{
11   "@context": "https://schema.org",
12   "@type": [
13     "SoftwareSourceCode",
14     "ComputationalWorkflow"
15   ],
16   "dct:conformsTo": "https://bioschemas.org/profiles/ComputationalWorkflow/1.0-RELEASE/",
17   "@id": "https://workflowhub.eu/workflows/1021?version=17",
18   "description": "<h1>\n \n \n \n <img alt=\"nf-core/scrnaseq\" src=\"docs/images/nf-core-scrnas
19   "name": "nf-core/scrnaseq",
20   "url": "https://workflowhub.eu/workflows/1021?version=17",
21   "keywords": "10x-genomics, 10xgenomics, alevin, bustools, Cellranger, kallisto, rna-seq, single-ce
22   "version": 17,
23   "license": "https://spdx.org/licenses/MIT",
24   "creator": [
25     {
26       "@type": "Person",
27       "@id": "#Peter%20J%20Bailey",
28       "name": "Peter J Bailey"
29     },
30     {
31       "@type": "Person",
32       "@id": "#Bailey%20PJ",
33       "name": "Bailey PJ"
34     },
35     {
36       "@type": "Person",
37       "@id": "#Alexander%20Peltzer",
38       "name": "Alexander Peltzer"
39     },
40     {
41       "@type": "Person",
42       "@id": "#Botvinnik%20O",
43       "name": "Botvinnik O"
44     },
45     {
46       "@type": "Person",
47       "@id": "#Olga%20Botvinnik",
48       "name": "Olga Botvinnik"
49     },
50   ],
51 }
```

|              |                                                                                                    |
|--------------|----------------------------------------------------------------------------------------------------|
| name         | nf-core/scrnaseq                                                                                   |
| url          | https://workflowhub.eu/workflows/1021?version=17                                                   |
| url          | https://workflowhub.eu/workflows/1021?version=17                                                   |
| keywords     | 10x-genomics, 10xgenomics, alevin, bustools, Cellranger, kallisto, rna-seq, single-cell, star-solo |
| version      | 17                                                                                                 |
| license      | https://spdx.org/licenses/MIT                                                                      |
| license      | https://spdx.org/licenses/MIT                                                                      |
| dateCreated  | 2025-03-26T09:57:52+00:00                                                                          |
| dateCreated  | 2025-03-26T09:57:52+00:00                                                                          |
| dateModified | 2025-03-26T09:57:53+00:00                                                                          |
| dateModified | 2025-03-26T09:57:53+00:00                                                                          |
| creator      |                                                                                                    |
| @type        | Person                                                                                             |
| @id          | https://workflowhub.eu/workflows/1021#Peter%20J%20Bailey                                           |
| name         | Peter J Bailey                                                                                     |
| creator      |                                                                                                    |
| @type        | Person                                                                                             |
| @id          | https://workflowhub.eu/workflows/1021#Bailey%20PJ                                                  |
| name         | Bailey PJ                                                                                          |
| creator      |                                                                                                    |
| @type        | Person                                                                                             |
| @id          | https://workflowhub.eu/workflows/1021#Alexander%20Peltzer                                          |
| name         | Alexander Peltzer                                                                                  |
| creator      |                                                                                                    |
| @type        | Person                                                                                             |
| @id          | https://workflowhub.eu/workflows/1021#Botvinnik%20O                                                |
| name         | Botvinnik O                                                                                        |

→ Search engines parse RDF metadata and better "understand" the content of the web page

## ② Inspecting metadata

# Approach



**Figure 1** Gathering, enriching and analyzing semantic web annotations in line with FAIR principles.

# schema.org



## Full Hierarchy

Schema.org is defined as two hierarchies: one for textual property values, and one for the things that they describe.

This is the main schema.org hierarchy: a collection of types (or "classes"), each of which has one or more parent types. Although a type may have more than one super-type, here we show each type in one branch of the tree only. There is also a parallel hierarchy for **data types**.

## Types:

Close hierarchy / Open hierarchy

- Thing
  - ▶ Action +
  - ▶ BioChemEntity +
  - ▶ CreativeWork +
  - ▶ Event +
  - ▶ Intangible +
  - ▶ MedicalEntity +
  - ▶ Organization +
  - ▶ Person +
  - ▶ Place +
  - Product
    - DietarySupplement
    - Drug
    - IndividualProduct
    - ProductCollection
    - ProductGroup

- ▶ General purpose **lightweight** ontology
- ▶ Aimed at **annotating web pages**
- ▶ Targetting **FINDABILITY**
- ▶ Originating from major search engines



# Schema.org is massively adopted

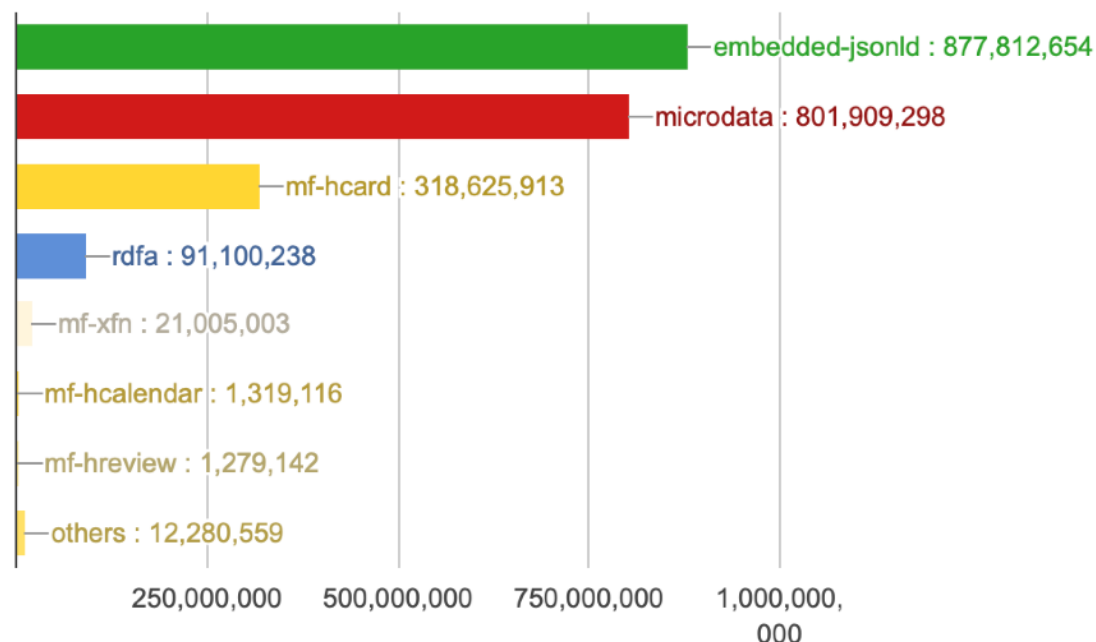
## Web Data Commons

Extracting Structured Data from the Common Crawl



|                        |                             |
|------------------------|-----------------------------|
| Crawl Date             | October 2022                |
| Total Data             | 82.71 Terabyte (compressed) |
| Parsed HTML URLs       | 3,048,746,652               |
| URLs with Triples      | 1,518,609,988               |
| Domains in Crawl       | 33,820,102                  |
| Domains with Triples   | 14,235,035                  |
| Typed Entities         | 19,072,628,514              |
| Triples                | 86,462,816,435              |
| Size of Extracted Data | 1.6 Terabyte (compressed)   |

### URLs with Triples



## Top Domains by Extracted Triples

1. [blogspot.com](https://www.blogspot.com) (879,564,145 triples)
2. [wordpress.com](https://www.wordpress.com) (458,770,038 triples)
3. [wikipedia.org](https://www.wikipedia.org) (190,087,065 triples)
4. [yummly.com](https://www.yummly.com) (87,112,540 triples)
5. [hotels.com](https://www.hotels.com) (81,991,039 triples)
6. [boohoo.com](https://www.boohoo.com) (79,884,394 triples)
7. [kayak.com](https://www.kayak.com) (77,623,248 triples)
8. [google.com](https://www.google.com) (73,729,078 triples)
9. [yahoo.com](https://www.yahoo.com) (65,317,838 triples)
10. [southleedslife.com](https://www.southleedslife.com) (63,758,451 triples)
11. [indiatimes.com](https://www.indiatimes.com) (58,899,559 triples)
12. [freepik.com](https://www.freepik.com) (56,124,447 triples)
13. [airbnb.com](https://www.airbnb.com) (51,964,983 triples)
14. [pinterest.com](https://www.pinterest.com) (47,251,484 triples)
15. [soundcloud.com](https://www.soundcloud.com) (45,745,317 triples)
16. [apple.com](https://www.apple.com) (42,410,414 triples)
17. [hostadvice.com](https://www.hostadvice.com) (42,309,867 triples)
18. [elpais.com](https://www.elpais.com) (42,136,136 triples)
19. [vsemayki.ru](https://www.vsemayki.ru) (38,167,517 triples)
20. [smugmug.com](https://www.smugmug.com) (38,031,434 triples)
21. [More](#)

Schema.org for  
Life-Science resources ?

# 37 ± Life Science profiles

| Name                                                                            | Group                                    | Use Cases                                                                           | Cross Walk                                                                            | Task & Issues                                                                         | Examples                                                                              | Live Deploys                                                                          |
|---------------------------------------------------------------------------------|------------------------------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| <u><a href="#">ChemicalSubstance</a></u><br>(v0.4-RELEASE)<br>07 April 2020     | <u><a href="#">Chemicals</a></u>         |    |    |    |    |    |
| <u><a href="#">ComputationalTool</a></u><br>(v1.0-RELEASE)<br>11 October 2021   | <u><a href="#">Tools</a></u>             |    |    |    |    |    |
| <u><a href="#">ComputationalWorkflow</a></u><br>(v1.0-RELEASE)<br>09 March 2021 | <u><a href="#">Workflow</a></u>          |    |    |    |    |    |
| <u><a href="#">DataCatalog</a></u><br>(v0.3-RELEASE-2019_07_01)<br>01 July 2019 | <u><a href="#">Data Repositories</a></u> |    |    |    |    |    |
| <u><a href="#">Dataset</a></u><br>(v0.3-RELEASE-2019_06_14)<br>14 June 2019     | <u><a href="#">Datasets</a></u>          |  |   |  |  |   |
| <u><a href="#">FormalParameter</a></u><br>(v1.0-RELEASE)<br>09 March 2021       | <u><a href="#">Workflow</a></u>          |  |  |  |  |  |
| <u><a href="#">Gene</a></u><br>(v1.0-RELEASE)<br>07 April 2021                  | <u><a href="#">Genes</a></u>             |  |  |  |  |  |
| <u><a href="#">MolecularEntity</a></u><br>(v0.5-RELEASE)<br>07 April 2020       | <u><a href="#">Chemicals</a></u>         |  |  |  |  |  |
| <u><a href="#">Protein</a></u><br>(v0.11-RELEASE)<br>07 April 2020              | <u><a href="#">Proteins</a></u>          |  |  |  |  |  |
| <u><a href="#">Sample</a></u><br>(v0.2-RELEASE-2018_11_10)<br>10 November 2018  | <u><a href="#">Samples</a></u>           |  |  |  |  |  |
| <u><a href="#">Taxon</a></u><br>(v0.6-RELEASE)<br>07 April 2020                 | <u><a href="#">Biodiversity</a></u>      |  |  |  |  |  |

- ▶ different usage of schema.org for life sciences
- ▶ Communities agree on **minimal**/**recommended**/**optional** annotation

# Metadata completeness

## R1.3: (Meta)data meet domain-relevant community standards

| Marginality: Recommended.                     |                                                                      |                                                                                                                                                                                                                                                                                                                                                                          |      |                                                                                                                                                                                             |
|-----------------------------------------------|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#"><u>applicationCategory</u></a>    | <a href="#"><u>Text</u></a><br><a href="#"><u>URL</u></a>            | <p><b>Schema:</b><br/>Type of software application, e.g. 'Game, Multimedia'.</p> <p><b>Bioschemas:</b><br/>Type of tool e.g. Command-line tool, Web application etc. <b>Note:</b> Bioschemas have changed <a href="#"><u>URL</u></a> to <a href="#"><u>Text</u></a> in the Expected Types. This will be reverted once Bio.Tools provides stable URIs for tool types.</p> | MANY | <p>Please use terms from the 'Tool type' table in the <a href="#"><u>biotools documentation</u></a></p>  |
| <a href="#"><u>applicationSubCategory</u></a> | <a href="#"><u>Text</u></a><br><a href="#"><u>URL</u></a>            | <p><b>Schema:</b><br/>Subcategory of the application, e.g. 'Arcade Game'.</p> <p><b>Bioschemas:</b><br/>Use an <a href="#"><u>EDAM:Topic</u></a> to describe the category of application</p>                                                                                                                                                                             | MANY | <p><a href="#"><u>EDAM:Topic</u></a></p>                                                                 |
| <a href="#"><u>author</u></a>                 | <a href="#"><u>Organization</u></a><br><a href="#"><u>Person</u></a> | <p><b>Schema:</b><br/>The author of this content or rating. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably.</p>                                                                                                                    | MANY |                                                                                                        |
| <a href="#"><u>citation</u></a>               | <a href="#"><u>CreativeWork</u></a><br><a href="#"><u>IRI</u></a>    | <p><b>Schema:</b><br/>A citation or reference to another creative work, such as another publication</p>                                                                                                                                                                                                                                                                  | MANY |                                                                                                        |

```
ex:myTool    rdf:type      schema:SoftwareApplication, prov:SoftwareAgent ;
schema:description "This tool does ... " ;
schema:license <https://spdx.org/licenses/MIT.html> ;
schema:codeRepository <http://github.com/...> .
```

# Profile → graph shapes

## Shapes Constraint Language (SHACL)

W3C Recommendation 20 July 2017

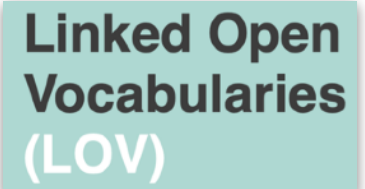
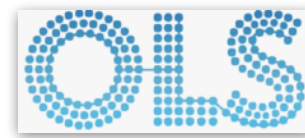


This version:

<https://www.w3.org/TR/2017/REC-shacl-20170720/>

```
1      ns:{{shape_name}} a sh:NodeShape ;
2      {% for c in target_classes %}
3      sh:targetClass {{c}} ;
4      {% endfor %}
5
6      {% for min_prop in min_props %}
7      sh:property [
8          sh:path {{min_prop}} ;
9          sh:minCount 1 ;
10         sh:severity sh:Violation
11     ] ;
12     {% endfor %}
13
14     {% for rec_prop in rec_props %}
15     sh:property [
16         sh:path {{rec_prop}} ;
17         sh:minCount 1 ;
18         sh:severity sh:Warning
19     ] ;
20     {% endfor %}
21     .
```

# Reuse of ontologies



F4: (Meta)data are registered or indexed in a searchable resource

I2: (Meta)data use vocabularies that follow the FAIR principles

R1.3: (Meta)data meet domain-relevant community standards

## Step 3: Metadata quality checks

Controlled vocabularies

Bioschemas

We now have a Knowledge Graph grounded to ontology concepts (classes) and relations (properties). Are these classes and properties already known in reference ontology registries such as [LOV](#), [OLS](#) or [BioPortal](#) ?

Check Vocabularies

Congratulations ! All Classes and Properties are referenced in one or more of the registries checked !

### Classes

|                                                                             |     |     |           |
|-----------------------------------------------------------------------------|-----|-----|-----------|
| <a href="http://schema.org/DataDownload">http://schema.org/DataDownload</a> | OLS | LOV | BioPortal |
| <a href="http://schema.org/Organization">http://schema.org/Organization</a> | OLS | LOV | BioPortal |
| <a href="http://schema.org/Person">http://schema.org/Person</a>             | OLS | LOV | BioPortal |
| <a href="https://schema.org/Dataset">https://schema.org/Dataset</a>         | OLS | LOV | BioPortal |

### Properties

|                                                                               |     |     |           |
|-------------------------------------------------------------------------------|-----|-----|-----------|
| <a href="http://ogp.me/ns#description">http://ogp.me/ns#description</a>       | OLS | LOV | BioPortal |
| <a href="http://ogp.me/ns#site_name">http://ogp.me/ns#site_name</a>           | OLS | LOV | BioPortal |
| <a href="http://ogp.me/ns#title">http://ogp.me/ns#title</a>                   | OLS | LOV | BioPortal |
| <a href="http://ogp.me/ns#url">http://ogp.me/ns#url</a>                       | OLS | LOV | BioPortal |
| <a href="http://schema.org/affiliation">http://schema.org/affiliation</a>     | OLS | LOV | BioPortal |
| <a href="http://schema.org/author">http://schema.org/author</a>               | OLS | LOV | BioPortal |
| <a href="http://schema.org/contentSize">http://schema.org/contentSize</a>     | OLS | LOV | BioPortal |
| <a href="http://schema.org/contentUrl">http://schema.org/contentUrl</a>       | OLS | LOV | BioPortal |
| <a href="http://schema.org/creator">http://schema.org/creator</a>             | OLS | LOV | BioPortal |
| <a href="http://schema.org/dateCreated">http://schema.org/dateCreated</a>     | OLS | LOV | BioPortal |
| <a href="http://schema.org/dateModified">http://schema.org/dateModified</a>   | OLS | LOV | BioPortal |
| <a href="http://schema.org/datePublished">http://schema.org/datePublished</a> | OLS | LOV | BioPortal |

# Metadata completeness

R1.3: (Meta)data meet domain-relevant community standards

Validation of Bioschemas profiles:

- **rank missing** metadata
- developer **focus** on **minimal** metadata first

Check BioSchemas

`https://workflowhub.eu/workflows/18?version=1` has type `http://schema.org/ComputationalWorkflow`

Using `https://bioschemas.org/profiles/ComputationalWorkflow/1.0-RELEASE` for validation, specified from the `dct:conformsTo` property.

| Required missing properties                                    | Improvements                                                                   |
|----------------------------------------------------------------|--------------------------------------------------------------------------------|
| <code>https://schema.org/input</code> <b>must be</b> provided  | <code>https://schema.org/citation</code> <b>should be</b> provided             |
| <code>https://schema.org/output</code> <b>must be</b> provided | <code>https://schema.org/contributor</code> <b>should be</b> provided          |
|                                                                | <code>https://schema.org/creativeWorkStatus</code> <b>should be</b> provided   |
|                                                                | <code>https://schema.org/documentation</code> <b>should be</b> provided        |
|                                                                | <code>https://schema.org/funding</code> <b>should be</b> provided              |
|                                                                | <code>https://schema.org/hasPart</code> <b>should be</b> provided              |
|                                                                | <code>https://schema.org/isBasedOn</code> <b>should be</b> provided            |
|                                                                | <code>https://schema.org/maintainer</code> <b>should be</b> provided           |
|                                                                | <code>https://schema.org/publisher</code> <b>should be</b> provided            |
|                                                                | <code>https://schema.org/runtimePlatform</code> <b>should be</b> provided      |
|                                                                | <code>https://schema.org/softwareRequirements</code> <b>should be</b> provided |
|                                                                | <code>https://schema.org/targetProduct</code> <b>should be</b> provided        |

# Hands-on session: improving metadata quality



## Question #6

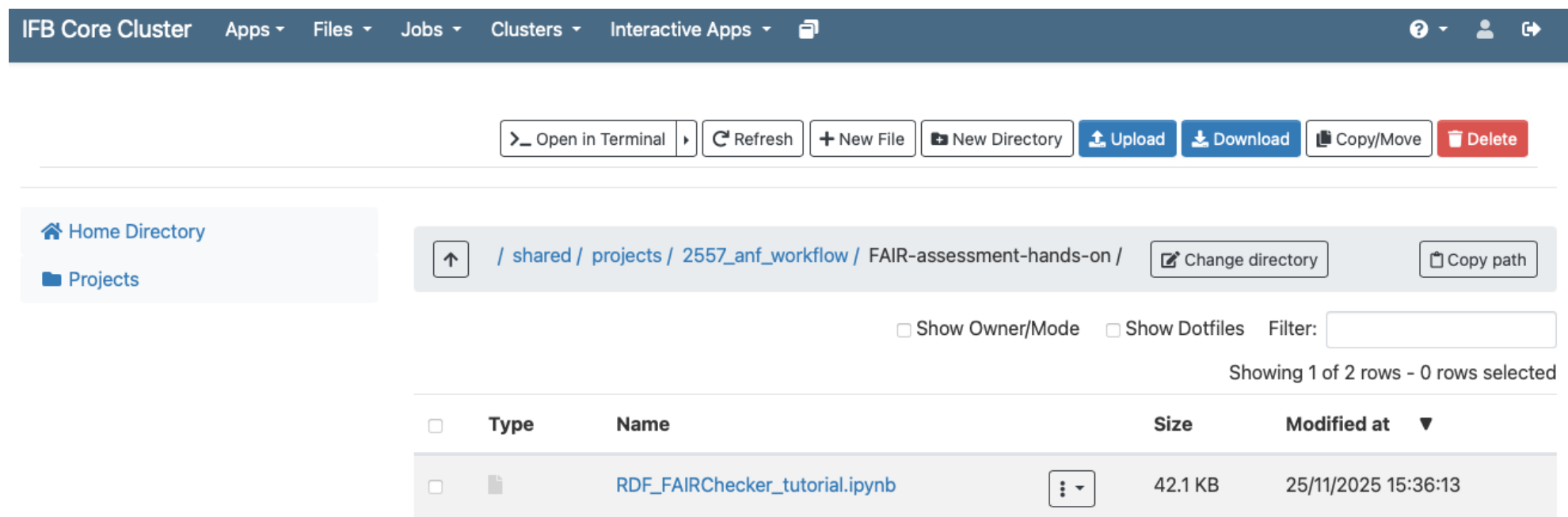
1. Use the FAIR-Checker "Inspect" module to identify **ontology term reuse**, and possibly **missing important metadata**
  - on <https://github.com/IFB-ElixirFr/fair-checker>  
<https://fair-checker.france-bioinformatique.fr>
  - on a zenodo dataset:  
<https://doi.org/10.5281/zenodo.17681549>
  - on a bioinformatics tool (Bio.Tools):  
<https://bio.tools/GrapeMine>
  - on a bioinformatics workflow (NF-Core):  
<https://nf-co.re/scrnaseq/4.1.0/>
  - on a bioinformatics workflow (WorkflowHub):  
<https://workflowhub.eu/workflows/1021>

2. Discuss the results

# Hands-on session: FAIR-Checker API

## Question #7

Run step-by-step the PART2 section of the Jupyter notebook available on the IFB cluster :



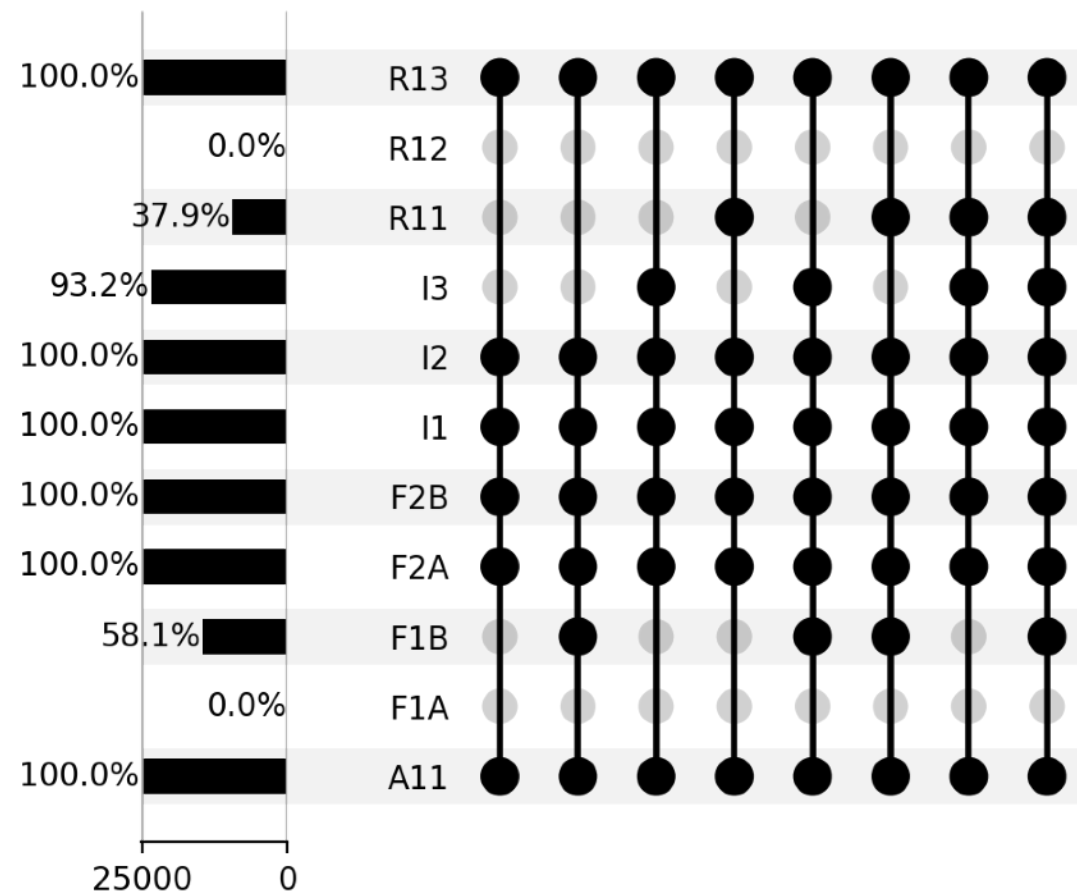
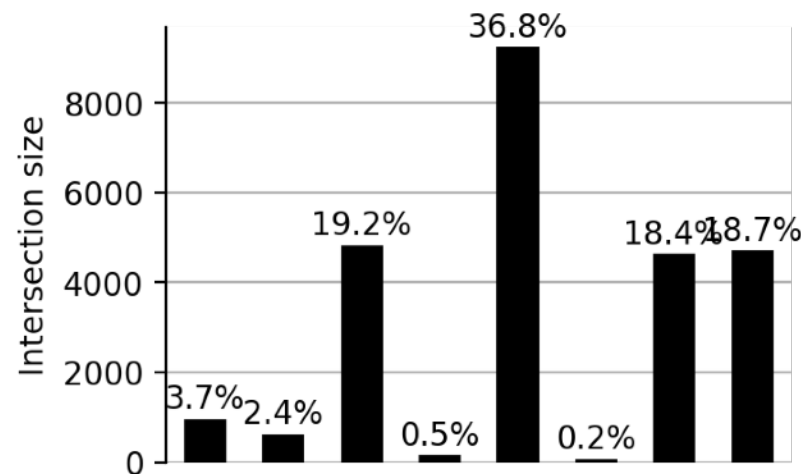
The screenshot shows the IFB Core Cluster file browser interface. The top navigation bar includes links for IFB Core Cluster, Apps, Files, Jobs, Clusters, and Interactive Apps. Below this, a toolbar contains buttons for Open in Terminal, Refresh, New File, New Directory, Upload, Download, Copy/Move, and Delete. The main area displays the current directory path: / shared / projects / 2557\_anf\_workflow / FAIR-assessment-hands-on /. A sidebar on the left shows the Home Directory and Projects. A table at the bottom lists files, with one file visible: RDF\_FAIRChecker\_tutorial.ipynb, which is 42.1 KB and was modified on 25/11/2025 at 15:36:13.

| Type | Name                           | Size    | Modified at         |
|------|--------------------------------|---------|---------------------|
| File | RDF_FAIRChecker_tutorial.ipynb | 42.1 KB | 25/11/2025 15:36:13 |

Using FAIR-Checker at  
large scale

# Large-scale FAIR metrics evaluations

How FAIR are Bio.Tools registered softwares ?

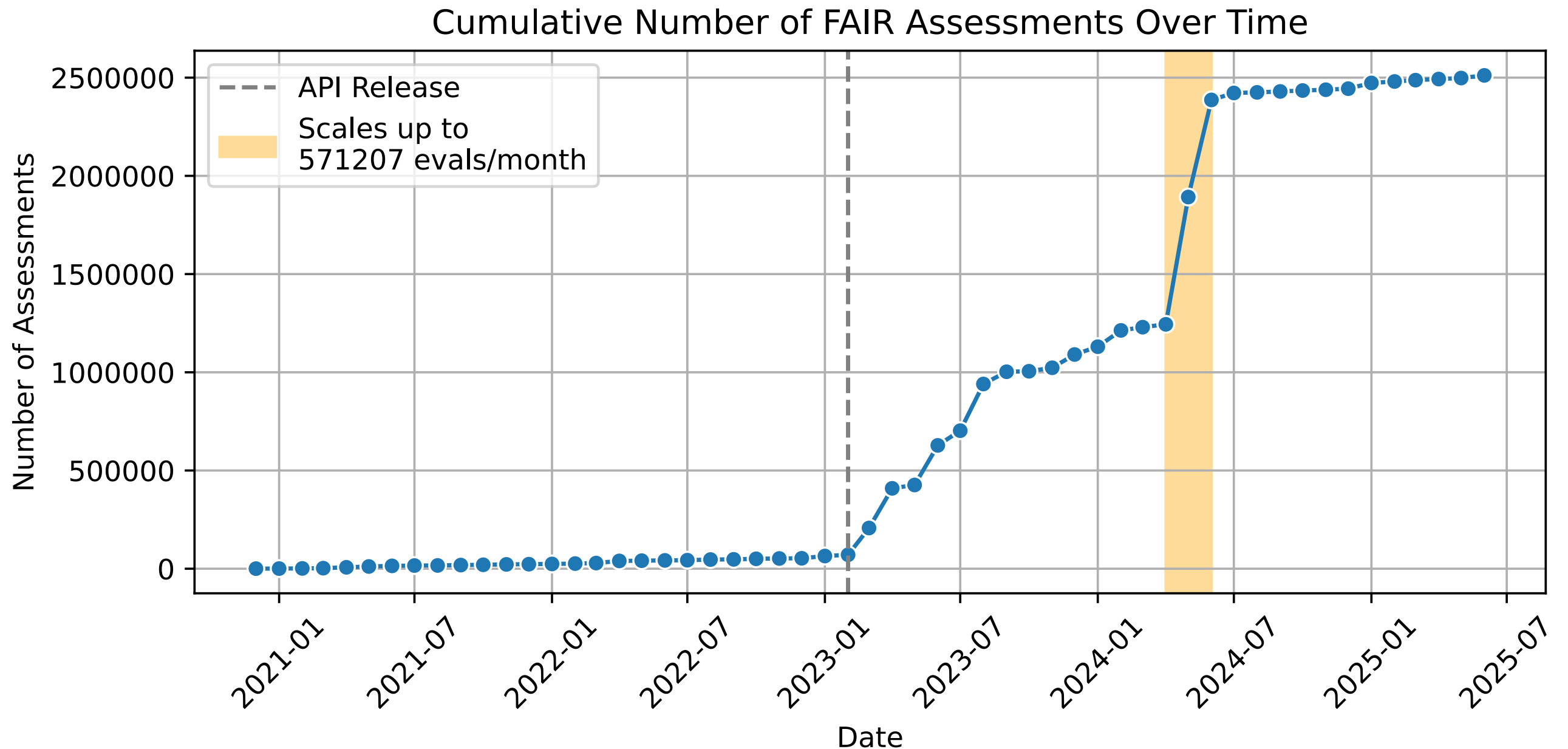


Running FAIR-Checker over  
**25.000+** bioinformatics softwares  
from Bio.tools.

R1.1: Only 37,9% of the tools  
expose a **licence**

R1.2: **No provenance** metadata  
→ massive impact if bio.tools  
developers provide PROV / PAV  
ontology terms

# Usage statistics

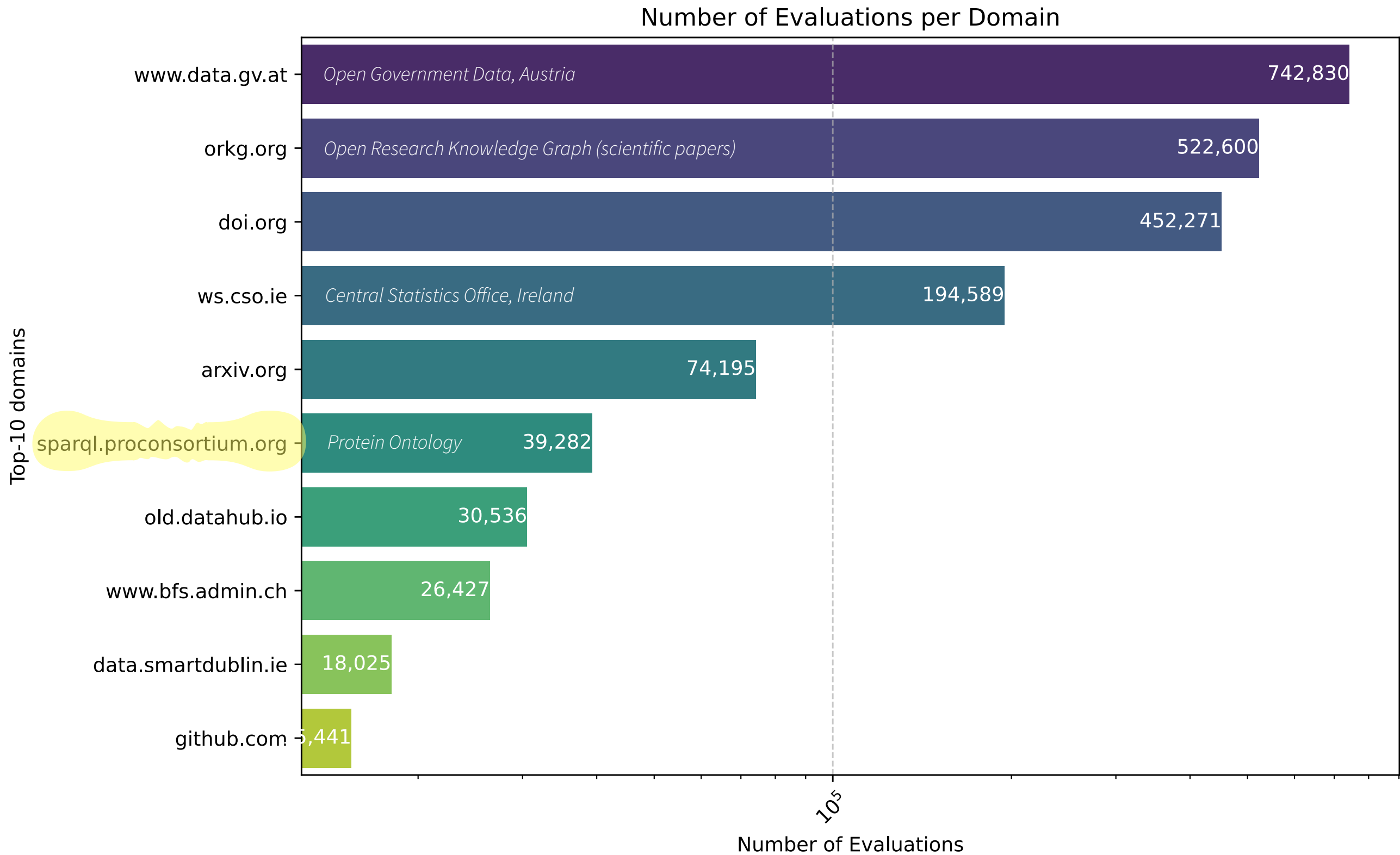


**102,180** unique target URLs

**47,537** unique target URLs evaluated more than 2 times

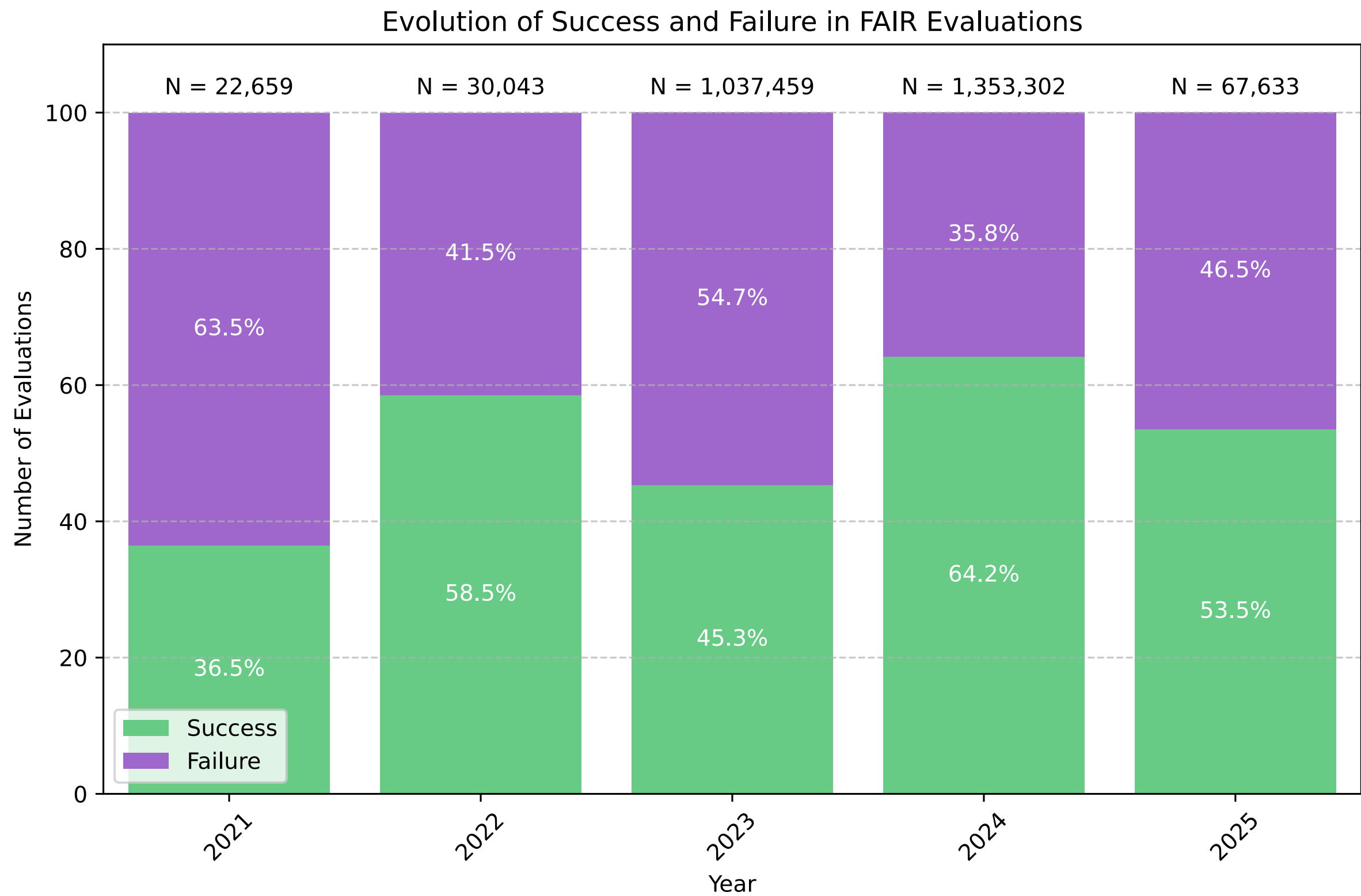
**437** unique target URLs evaluated more than 10 times

# Most evaluated domains



# Evolution of FAIR scores

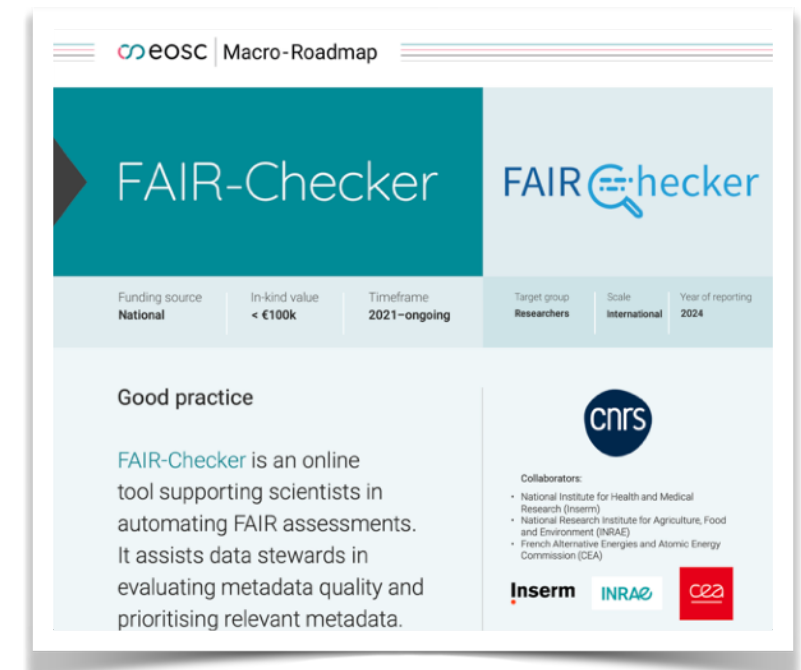
Work in progress



# A community service, with possible improvements



<https://fair-checker.france-bioinformatique.fr>



## Future works

- ▶ Support "FAIR-Signposting" for **better metadata consumption**
- ▶ Bioschemas **profile recommender**
- ▶ Allow users to provide **missing metadata**
- ▶ Suggest semantic metadata based on **AI generation** pipeline
- ▶ Retrospective **usage study**



Recap'

# A practical overview on FAIR assessment, in 2 hours ...

- ▶  **Quick intro on FAIR principles**
- ▶ **Knowledge graphs and semantic metadata**
  -  What are Knowledge Graphs, Computational Ontologies ?
  -  How to create Knowledge Graphs ? (team work, "pen & paper")
  -  How to write machine-readable semantic metadata ? (demo, python code)
  -  How to query knowledge graphs with SPARQL (demo, python code)
- ▶ **Tools and resources for increased FAIRness**
  -  FAIR-Checker
  -  Tools and workflow registries
  -  Using FAIR-Checker and interpreting the FAIR assessment results (web tool)
  -  Using the FAIR-Checker API and processing multiple resources (python code)

# Additional resources

- ▶ W3C RDF Primer (<https://www.w3.org/TR/rdf11-primer/>)
- ▶ W3C JSON-LD Primer (<https://json-ld.org/primer/latest/>)
- ▶ JSON-LD playground (<https://json-ld.org/playground/>)
- ▶ [schema.org](https://validator.schema.org/) validator (<https://validator.schema.org/>)
- ▶ Gaignard, A., Rosnet, T., de Lamotte, F., Lefort, V., & Devignes, M. (2023). *FAIR-Checker: supporting digital resource findability and reuse with Knowledge Graphs and Semantic Web standards*. Journal of Biomedical Semantics, 14. <https://doi.org/10.1186/s13326-023-00289-5>
- ▶ Lamarre, P., Andersen, J., Gaignard, A., Cazalens, S. (2025). *A Deep Dive into FAIRness Assessment: UReFM, a Formal Framework for Representing, Analyzing and Comparing Measures*. In: Hameurlain, A., Tjoa, A.M. (eds) Transactions on Large-Scale Data- and Knowledge-Centered Systems LVIII. Lecture Notes in Computer Science(), vol 16080. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-662-72116-2\\_4](https://doi.org/10.1007/978-3-662-72116-2_4)

# Acknowledgements



Marie-Dominique Devignes



Thomas Rosnet



Frédéric de Lamotte

## *Funding:*

French institute for Bioinformatics (Elixir-FR)

contact

[alban.gaignard@univ-nantes.fr](mailto:alban.gaignard@univ-nantes.fr)

Back-up slides

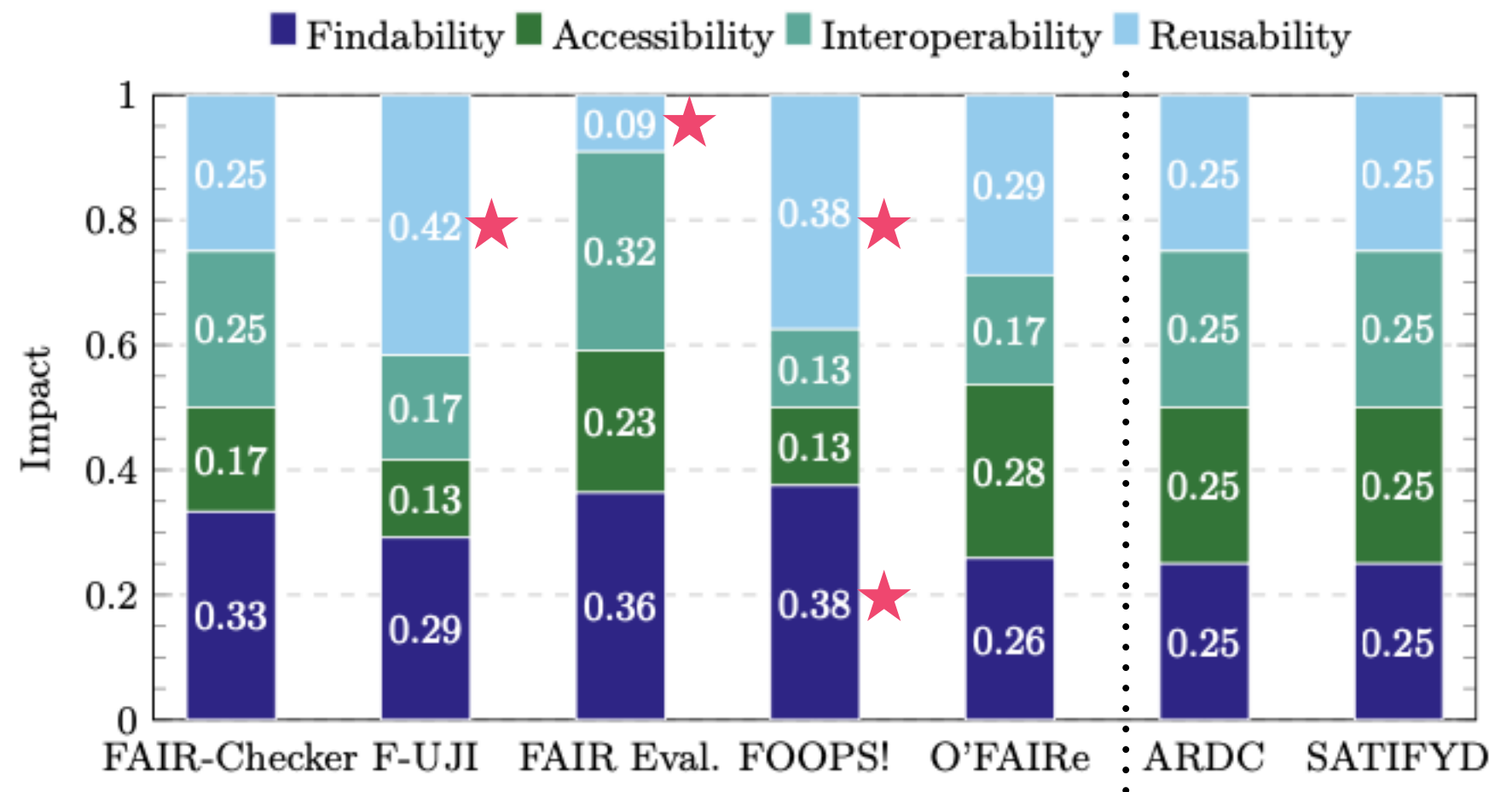
# Do engines reach consensus on FAIR assessment ?

| Resource                                   | F-UJI (%) |                                                                                      | FAIR-Checker (%) | Std dev |
|--------------------------------------------|-----------|--------------------------------------------------------------------------------------|------------------|---------|
| Dataset (PANGAEA) [31]                     | 91        |                                                                                      | 91.70            | 0.49    |
| Gene Ontology (OLS) [21]                   | 18        |   | 16.70            | 0.92    |
| Dataset (Harvard Dataverse) [23]           | 75        |                                                                                      | 79.20            | 2.97    |
| Dataset (Kaggle) [26]                      | 60        |                                                                                      | 70.80            | 7.64    |
| Online course (Moodle) [28]                | 4         |                                                                                      | 16.70            | 8.98    |
| ..... Dataset (Governmental platform) [22] | 52        |                                                                                      | 70.80            | 13.29   |
| Dataset (WHO) [39]                         | 27        |  | 50.00            | 16.26   |
| Training material (TeSS) [36]              | 39        |                                                                                      | 70.80            | 22.49   |
| Bioinformatics tool (bio.tools) [6]        | 18        |                                                                                      | 54.20            | 25.60   |
| Dataset (RDF metadata) [33]                | 43        |                                                                                      | 87.50            | 31.47   |

- ▶ Higher scores for FAIR-Checker
- ▶ Last two entries: std. dev. > 25 % ?

# How much biased are FAIR assessment tools ?

*Are all principle equally contributing to the global FAIR assessment score ?*



→ *How to get a good FAIR score with a minimal effort ?*

- ▶ Pay attention to identifiers (F), license + provenance + domain-specific standards (R) if you use FOOPS!
- ▶ Not useful to spend energy on provenance or domain ontologies if you use FAIR-Evaluator ...
- ▶ ... but pay attention to it if you use F-UJI.

# LLM-based generation of RDF triples



**You**

Can you translate in RDF (turtle syntax) the following statement : "The insulin receptor (IR) is a transmembrane receptor that is activated by insulin, IGF-I, IGF-II and belongs to the large class of receptor tyrosine kinase"



**ChatGPT**

Sure, here's how you could represent that statement in RDF using Turtle syntax:

turtle

Copy code

```
@prefix ex: <http://example.org/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .

ex:InsulinReceptor rdf:type ex:TransmembraneReceptor ;
    ex:activatedBy ex:Insulin,
                  ex:IGF-I,
                  ex:IGF-II ;
    ex:belongsToClass ex:ReceptorTyrosineKinase .

ex:Insulin a ex:Hormone .
ex:IGF-I a ex:Hormone .
ex:IGF-II a ex:Hormone .
```



# Hallucinated LLM-based interpretation of KGs



La protéine P06213 est une protéine. Elle est citée dans des journaux ayant pour titre "The human insulin receptor cDNA: the structural basis for hormone-activated transmembrane signaling." Ses auteurs sont Ebina Y. et Ellis L. Cette protéine a des annotations incluant "Sous-unité alpha du récepteur à l'insuline" avec une masse de 83642. Cette annotation correspond à une région spécifique identifiée aux positions 28 à 758.

!! inaccurate positions

!! no link to the pubmed paper